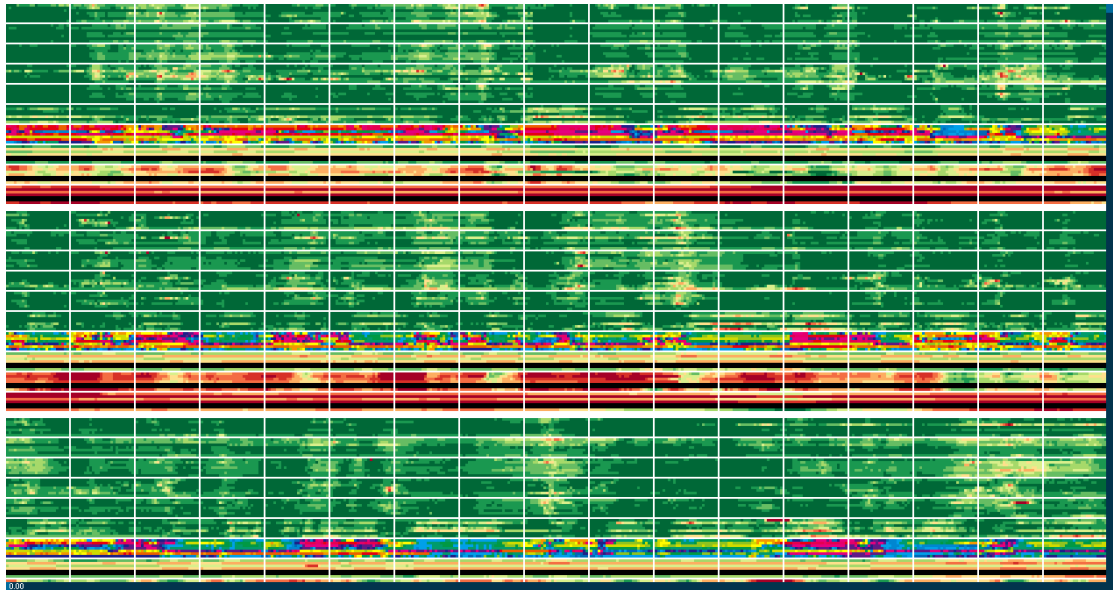


本周周报（2013.10.8-2013.10.13）

郭方舟

本周工作

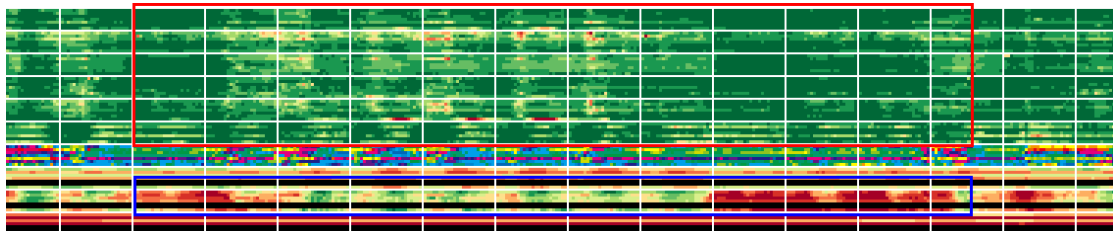
1. 空气污染数据可视化



本周将青岛市空气数据重构后导入到 MySQL 数据库中，并使用像素图的方法对数据进行了初步可视化。

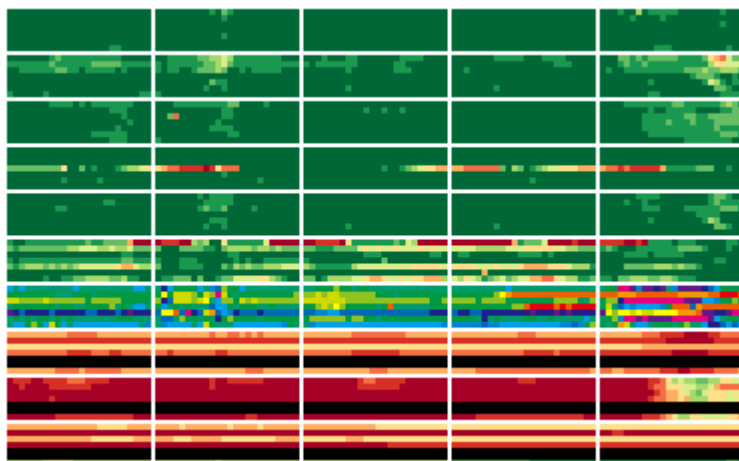
空气数据有十个属性，分别是 CO，NO2，PM10，SO2，NOX，WS，WD，TEMP，HUMD，PRESS，共有 7 个观测点。在可视化中，数据分层为月份-属性-观测点，横轴为时间轴。为了区分日期和月份，在横向上以一天为单位进行了分割，在纵向上以一个月为单位进行分割。

在目前的可视化结果上已经可以看出一些有意思的东西，比如：



蓝色框表示的是湿度，红色框表示的 CO 等五个代表空气污染的属性，从图中可以看到，污染与空气湿度有着明显的联系，空气湿度增大后空气质量有明显的改善。

同时在可视化结果中也能容易的看到一些明显的异常点，比如：



这里的 SO_2 指标出现了明显的异常，但是进一步的原因还有待查明，我们怀疑是有烟花燃放的时间，但是还没有找到很好的证据。

因为下一步工作需要计算相关性，所以看了 CCA 算法的介绍，并在网上找到了一个可用的 python 库，这个库集成了很多机器学习的算法。下周准备使用 python 计算数据点之间的相关性。

周日讨论后定下进一步的工作，包括：

- 1) 为了突出事件，找出一种自适应的方法将事件标识出来，信息熵是下周要进行尝试的方法。
- 2) 添加用户对颜色映射的选择，通过不同的颜色映射，可能可以表达出数据中不同的 pattern。
- 3) 顺序调整
通过调整显示顺序，使原本隐含的时间暴露出来。

2. 整理 ftp 并生成目录

这个工作进行了一部分，但是有一些文件夹还需要问一下小马哥要怎么处理。

3. 读《Visualization Design and Analysis: Abstractions, Principles, and Methods》和《Grammar of Graphics》

开始读这两本书，但是由于主要在做空气污染数据，所以进度不大，都只看到第一章。下周将这两本书的前两章读完。

下周工作

1. 继续进行空气污染数据的可视化工作。
2. 读书